Available online at
www.ijacskros.com

Indian
Journal of Advances in
Chemical Science

Indian Journal of Advances in Chemical Science 4(4) (2016) 374-385

Antifibrosis Two-dimensional and Three-dimensional Quantitative Structure-activity Relationship Studies on N_1-substituted Phenylhydroquinolinone Derivatives

M. Abul Kashem Liton*, M. Idrish Ali

Department of Chemistry, Mawlana Bhashani Science and Technology University, Santosh, Tangail - 1902, Bangladesh

Received 24th July 2016; Revised 18th September 2016; Accepted 19th September 2016

ABSTRACT

A series of 35 N_1-substituted phenylhydroquinolinone derivatives were subjected to two-dimensional (2D)- and three-dimensional quantitative structure-activity relationship (3D-QSAR) studies applying multiple linear regression (MLR), principal component regression (PCR), partial least square regression (PLSR), artificial neural network (ANN), and molecular interaction energy fields (MIFs) analyses. The best predictive model obtained by MLR, PCR, PLSR, ANN for 2D-QSAR, and MIFs at 2.0 Å grids spacing for 3D-QSAR gave the cross-validated correlation coefficient q^2 of 0.589, 0.538, 0.534, 0.815, and 0.590 and squared correlation coefficient r^2 of 0.863, 0.848, 0.791, 0.954, and 0.940, respectively. The statistically significant models were established from a training set of 28 molecules, which were validated by evaluation of test set of 7 molecules. The predicted antifibrosis activities showed a very good agreement with experimental values in both 2D- and 3D-QSAR studies.

Key words: *Multiple linear regression, Principal component regression, Partial least square regression, Artificial neural network, Molecular interaction energy fields, Leave one out.*

1. INTRODUCTION

Fibrosis is the formation of excess fibrous connective tissue in an organ or tissue in a reparative process. The causes of fibrosis can be radiation, chemotherapy, burns, or an improper treatment of lymphedema. This can be a reactive, benign, or pathological state. In response to injury, this is called scarring and if fibrosis arises from a single-cell line, this is called a fibroma. Physiologically, this acts to deposit connective tissue, which can obliterate the architecture and function of the underlying organ or tissue. Fibrosis can be used to describe the pathological state of excess deposition of fibrous tissue, as well as the process of connective tissue deposition in healing [1]. Fibrosis is similar in the process of scarring, in that both involve stimulated cells laying down connective tissue, including collagen and glycosaminoglycans. Immune cells called macrophages, and damaged tissue between surfaces called interstitial release transforming growth factor-β (TGF-β). This can be occurred due to the numerous reasons, including inflammation of the nearby tissue, or a generalized inflammatory state, with increased circulating mediators. TGF-β stimulates the proliferation and activation of fibroblasts, which deposit connective tissue [2]. Currently, treatments for fibrotic diseases such as idiopathic pulmonary fibrosis [3], liver cirrhosis, systemic sclerosis, progressive kidney disease, and cardiovascular fibrosis typically target the inflammatory response [4]. Pirfenidone (5-methyl-1-phenyl-2-(1H)-pyridone, PFD) is orally active approved drug for the treatment of fibrosis diseases and has antifibrotic activity in animal models of lung, heart, kidney, liver, and pulmonary fibrosis [5]. *In vitro*, this compound inhibits the synthesis of growth factor beta (TGF-β) and decreases the extracellular matrix, and blocks the mitogenic effect of profibrotic cytokines in adult human lung fibroblasts derived from patients with idiopathic pulmonary fibrosis [6]. Bleomycin is an anticancer agent prescribed in the treatment of cancers such as a lung. At the stage of bleomycin-induced lung damage, several biochemical and functional changes occur, such as inflammatory cell infiltration, increased collagen content, and reduced lung volume [7-9]. Despite their proven efficacy, one major shortcoming of PFD is the fact that their activity is not high; therefore, they require an elevated dosage [10] to be effective. The research group, Yuan *et al*. had synthesized fluorofenidone [11], which showed lower toxicity but still equivalent activity to PFD. Therefore, there is a need for an alternative drug that can overcome the shortcomings of the PFD.

*Corresponding Author:
E-mail: litonchem@hotmail.com
Phone: +88-0921-62243

Now, the effect of antifibrotic agents and its potential mechanism are frequently studied. The purpose of the present study is to investigate the quantitative structure-activity relationships (QSARs) of some antifibrotic agents through physiochemical properties of molecular structures and molecular interaction energy fields (MIFs) as well as comparative studies among the proposed models in QSAR.

2. EXPERIMENTAL

2.1. Biological Assay

The experimental antifibrosis activity [12] of the target compounds 1-35 was obtained from Ling *et al*. The derivatives were synthesized from a precursor compound (Table 1) using various substituents R, and the antifibrosis activity of the target compounds (1-35) was evaluated on the NIH3T3 cell line using MTT methods.

2.2. Data Set

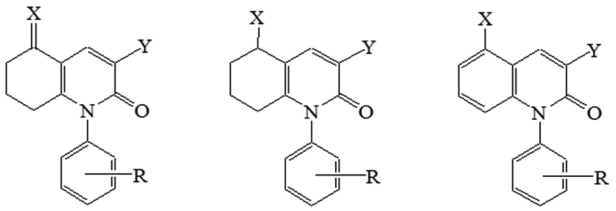
2D- and 3D-QSAR modelling were applied to a set of 35 molecules, which were divided into a training set [13] of 28 molecules, and a test set of 7 molecules in a random manner (the test set is marked by *). According to research methodology, all experimental IC_{50} values (mM) were converted to the tenth-based logarithm of IC_{50}, i.e., $LogIC_{50}$ and used as the dependent variable in QSAR studies. The structures of these compounds and their experimental biological activities are shown in Table 1.

3. COMPUTATIONAL DETAILS

3.1. Structure Generation and Optimization

At first, all the 3D-structures were generated by Gauss View 03 software, and minimization of 3D-structure was performed by the MOPAC-2012 software [14] using the semi-empirical (PM6) method [15]. All geometric variables were finally optimized for each compound using a Gaussian03W program (v6.0) [16] at the level of B3LYP/6-31G(d,p) theory [17] and the low-energy conformers were ensured with all real frequencies by the frequency calculations, and these lowest energy conformations were employed in QSAR analysis.

Table 1: The structure and *in vitro* MTT assay results on NIH3T3 cell line for compounds 1-35.



1-25 26-34 35

ID No.	R	X	Y	IC_{50} (Exp)	$LogIC_{50}$ (Exp)	ID No.	R	X	Y	IC_{50} (Exp)	$LogIC_{50}$ (Exp)
1	H	O	COOH	1.0	0.00	19	p-F	O	H	4.0	0.60
2	o-Cl	O	COOH	20.0	1.30	20	p-Br	O	H	2.2	0.34
3	p-Cl	O	COOH	2.0	0.30	21	p-OCH_3	O	H	2.5	0.40
4	p-CH_3	O	COOH	18.7	1.27	22	2,4-di-Cl	O	H	6.2	0.79
5	p-F	O	COOH	2.6	0.41	23	2,4-di-CH_3	O	H	3.7	0.57
6	p-Br	O	COOH	4.2	0.62	24	p-F	N-OH	H	3.9	0.59
7	2,4-di-Cl	O	COOH	25.6	1.41	25	p-OCH_3	N-OH	H	83.0	1.92
8	2,4-di-CH_3	O	COOH	15.6	1.19	26	H	OH	H	6.5	0.81
9	3,4-di-Cl	O	COOH	0.3	−0.52	27	p-Cl	OH	H	1.5	0.18
10	2-Cl, 4-F	O	COOH	2.7	0.43	28	o-F	OH	H	1.0	0.00
11	H	O	H	5.4	0.73	29	p-F	OH	H	2.7	0.43
12	o-Cl	O	H	3.5	0.54	30	p-Br	OH	H	0.3	−0.52
13	m-Cl	O	H	4.1	0.61	31	p-OCH_3	OH	H	3.4	0.53
14	p-Cl	O	H	2.8	0.45	32	2,4-di-Cl	OH	H	1.0	0.00
15	o-CH_3	O	H	3.5	0.54	33	2,4-di-CH_3	OH	H	1.7	0.23
16	m-CH_3	O	H	4.0	0.60	34	H	NH_2	H	2.1	0.32
17	p-CH_3	O	H	2.4	0.38	35	2-Cl, 4-F	OH	COOH	0.09	−1.05
18	o-F	O	H	2.2	0.34						

IC_{50} values are in mM (mill mole). The tenth-based $LogIC_{50}$

3.2. Molecular Descriptor Calculation

Molecular descriptors were calculated after minimization of the energy of the data set molecules. Briefly, the Padel Descriptor software (v2.12) [18] was employed to calculate the molecular descriptors. Overall, more than 480 theoretical descriptors were calculated. These descriptors can be classified into several groups, including: (I) Constitutional, (II) geometrical, (III) topological, (IV) electronic (V) Burden–CAS-University of Texas, and (VI) weighted holistic invariant molecular (WHIM). Quantum chemical (QM) descriptors such as HOMO and LUMO energies, heat of formation, dipole moment, square dipole moment, surface area, surface volume, energy gap, and total energy were further calculated using Gaussian03W (v6.0) program at the level of B3LYP/6-31G (d,p) theory. From 493 different descriptors, including quantum chemical which having <0.8 correlations was retained for further analyses, and finally, the best 2D-QSAR models were obtained (optimum r^2 and q^2 values with the least number of descriptors) with the referred descriptors which are shown in Table 2 with a detailed description.

3.3. Alignment

The molecules were superimposed using atom-based alignment by the open3DALIGN tool [19,20] given in Figure 1. From the data set, the compound 35 shown in

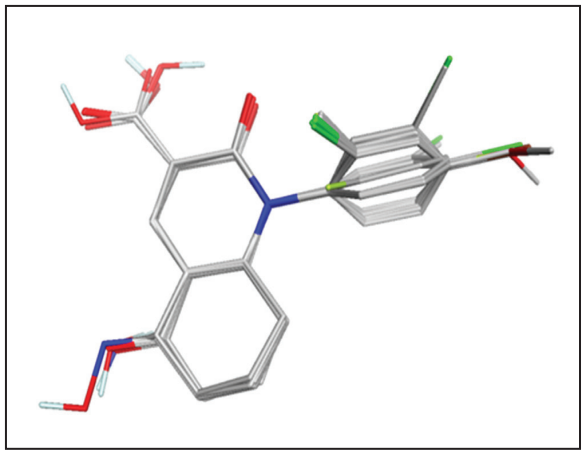


Figure 1: The atom-based alignment of the superimposed structure of all compounds used in the three-dimensional quantitative structure-activity relationship (molecular interaction energy fields) analysis.

Table 2: Molecular descriptors used in 2D-QSAR study.

Type	Description
ATES descriptor	
LipoAffI	LipoAffinity Index [24] is obtained by the equation of $LA=\Sigma(a_iS_i)$ where i≠N and O; a_i is the LR coefficient, and S_i is the E-state index
ETA descriptor	
ETA_EtaP	Composite index Eta relative to molecular size
CPSA descriptors	
WNSA2	Surface weighted charged partial negative charged surface area
RPCS	Relative positive charged surface area
WHIM descriptors	
WV_{unity}	WV_{unity} is a global WHIM descriptor of $V=\lambda_1+\lambda_2+\lambda_3+\lambda_1\lambda2+\lambda_2\lambda_3+\lambda_3\lambda_1+\lambda_1\lambda_2\lambda_3$ weighted by a suffix-unity
WV_{volume}	WV_{volume} is a global WHIM descriptor of $V=\lambda_1+\lambda_2+\lambda_3+\lambda_1\lambda2+\lambda_2\lambda_3+\lambda_3\lambda_1+\lambda_1\lambda_2\lambda_3$ weighted by a suffix-volume (van der Waals volumes)
WV_{eneg}	WV_{eneg} is a global WHIM descriptor of $V=\lambda_1+\lambda_2+\lambda_3+\lambda_1\lambda2+\lambda_2\lambda_3+\lambda_3\lambda_1+\lambda_1\lambda_2\lambda_3$ weighted by a suffix-eneg (Mulliken atomic electronegativities)
$Wnu2_{unity}$	$Wnu2_{unity}$ is a directional WHIM descriptor of $nu_m=\frac{\lambda^2 A}{\Sigma_i t_i^4}$ where m=2 weighted by a suffix-unity
$Wlamda1_{unity}$	$Wlamda1_{unity}$ is a directional WHIM descriptor of λ_m where m=1 weighted by a suffix-unity
Quantum chemical (QM) Descriptors	
E_{HOMO}	Highest occupied molecular orbital energy
E_{LUMO}	Lowest unoccupied molecular orbital energy
ΔH_{FORM}	Heat of formation
ΔE_{GAP}	Energy difference between E_{HOMO} and E_{LUMO} energies

WHIM: Weighted holistic invariant molecular, CPSA: Charged partial surface area, ETA: Extended topochemical atom, ATES: Atom type electrotopological state

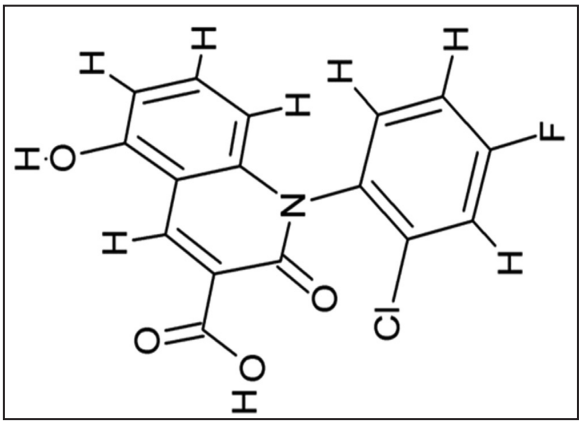


Figure 2: The most active compound, 35 used as a template for alignment of data set.

Figure 2 was selected due to its high biological activity to construct the template as a representative structure of other compounds, and the atom-based alignment was done by the Open3DALIGN tool. Except for some special notes; default values were chosen, and their geometries were optimized by the RMS gradient [21] criterion method of MMFF94s force-field implemented in the TINKER software program [22] using command options from open3DALIGN tool. The energy convergence criterion is 0.01 kcal/mol.

3.4. MIFs Calculation

The MIFs are the interaction energies between a probe atom (or a molecule) and a set of aligned molecules. To generate the MIFs, a probe atom is systematically moved from one point to another for each aligned molecule within 3D grid spacing [23]. At each grid point, the interaction energy is calculated between the probe and the target molecule. In this study, the 35 aligned molecules were placed in 2.0 Å 3D cubic lattice spacing. The steric (van der Waals) and electrostatic (Coulombic) interaction energies were calculated for each molecule at that grid point using a "CR" Alkyl Carbon probe (default) with automatically assigned charges (+1) using OpenBabel utilities.

4. VALIDATION OF THE MODELS

The models were validated internally (leave-one-out method, [LOO]) and externally. In internal validation [24,25], a compound is eliminated from the training set and its biological activity is calculated by the model derived from the rest of the compound in that training set. This step is repeated until every compound in the training set has been eliminated, and its activity is predicted. The cross-validation [26], q^2 was determined using the following equation where y_i and $\hat{y}_i$ are actual and predicted values of i^{th} compound in the training set, respectively, and y_{mean} is the actual average activity of all compounds in the training set.

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2} \quad (1)$$

External validation r^2_{pred} is also done by calculating the correlation coefficient r^2_{pred} value using Equation 3, where y_i and $\hat{y}_i$ are the actual and predicted activities of the i^{th} compound in test sets, respectively. The q^2 value represents the predictive ability of the proposed model, and the obtained r^2_{pred} value indicates the predictive power of the QSAR model for the external test set. The cross-validated (internal validation) analysis was also selected on the highest value of q^2 and lowest value of SPRESS in the training set.

5. RESULTS AND DISCUSSIONS

5.1. Development of the 2D-QSAR Models

For the 2D-QSAR study, model selection was performed by BuildQSAR (ver. 2.1.0.0) software program [27]. Statistical calculations allowed for the selection of the models with the following characteristics: High squared correlation coefficient (r^2), high Fischer's value (F test), low standard error of calculation (SDEC), and the least number of descriptors involved. Next, the IBM SPSS Statistics20 software package [28] was applied for detailed statistical analysis of the 2D-QSAR models. Only those descriptors having an intercorrelation coefficient <0.80 were considered for the present study described in Table 2. The generated best model was validated for the predictive ability internally (LOO on the training set) and externally (test set). Statistically, the best significant 2D-QSAR models on the minimum SDEC by multiple linear regression (MLR), principal component regression (PCR), partial least square regression (PLSR), and artificial neural network (ANN) regression methods were selected for discussions.

MLR (Model 1)

$$\log IC_{50} = 13.9877 \times Wnu2_{unity} - 7.1107 \times ETA\_Etap + 1.0345 \times " E_{GAP} + 0.2949 \times RPCS + 0.1060 \times WV_{volume} + 0.01962 \times WNSA2 - 10.0552 \quad (2)$$

$n_{train} = 28; r^2 = 0.863 SDEC = 0.205; Ftest = 21.089;$
$p < 0.0001; q^2 = 0.589; SPRESS = 0.315;$
$n_{test} = 7; r^2_{pred} = 0.801; SDEP = 0.236$

PCR (Model 2)

$$\log IC_{50} = 7.0942 \times Wnu2_{unity} - 0.5719 \times LipoAffI + 0.4230 \times E_{HOMO} + 0.2397 \times RPCS + 0.01667 \times WNSA2 + 0.0377 \times WV_{eneg} - 2.7537 \quad (3)$$

$n_{train} = 28; r^2 = 0.848; SDEC = 0.216; Ftest = 18.619;$
$p < 0.0001; q^2 = 0.538; SPRESS = 0.349;$
$n_{test} = 7; r^2_{pred} = 0.694; SDEP = 0.293$

PLSR (Model 3)

$$\log IC_{50} = 1.090 \times Wlambda1_{unity} - 1.088 \times LipoAffI + 0.195 \times RPCS + 0.023 \times WV_{eneg} + 0.0111 \times WNSA2 + 0.001 \times \Delta H_{FORM} + 5.198 \quad (4)$$

$n_{train}=28; r^2=0.791; SDEC=0.254; Ftest=12.60;$
$p<0.0001; q^2=0.534; SPRESS=0.349;$
$n_{test}=7; r^2_{pred}=0.644; SDEP=0.316$

ANN (Model 4)

$n_{train}=28; r^2=0.954; SDEC=0.119; Ftest=67.168;$
$p<0.0001; q^2=0.815; SPRESS=0.215;$
$n_{test}=7; r^2_{pred}=0.862; SDEP=0.196$

For ANN model, it is true that the ANN can be considered as a nonlinear model [29,30], and it has an intrinsic ability to incorporate nonlinear dependencies between the dependent and independent variables ($Wnu2_{unity}$, WV_{unity}, WNSA2, RPCS, LipoAffI, and ΔH_{FORM}) without using an explicit mathematical equation.

In the above QSAR models, r^2 is a squared correlation coefficient that has been multiplied by 100 gives explained variance in biological activity. The predictive ability of generating QSAR models was evaluated by q^2 employing LOO method. F value reflects the ratio of variance explained by models and variance due to an error in regression. The high F value indicates that the model is robust and statistically significant. Low SDEC of calculation suggested that models are statistically significant. The predictive ability of the QSAR model was also confirmed by external validation of test set compounds denoted by r^2_{pred}. Among these four models, ANN has come out with very good results as compare to other three models. Results of ANN analysis showed very good predictive ability as indicated by q^2, SPRESS, F-test, and r^2_{pred} values (Table 3).

5.2. Evaluation of 2D-QSAR Models

The data used in these experiments consisted of 35 N_1-phenylhydroquinolinone derivatives, and antifibrosis activity was calculated by $LogIC_{50}$. This dataset was divided into a training set of 28 molecules of developing the MLR, PCR, PLSR, and ANN models and a test set of 7 compounds for evaluating the predictive ability of the models. The training and test set were same for all regression methods. The experimental and predicted $LogIC_{50}$ values of all molecules studied by MLR, PCR, PLSR, and ANN in this work were shown in Table 4. The predictive model building abilities of four methods, MLR, PCR, PLSR, and ANN were compared.

MLR (Model 1), PCR (Model 2), and PLSR (Model 3) show the positive contributions of WHIM [31], CPSA [32], and QM [33-35] descriptors and negative contributions of ATES [36,37] and ETA [38-40] descriptors which are shown in Table 2 with detailed description. WHIM descriptors ($Wnu2_{unity}$, WV_{volume}, WV_{eneg}, $WLambda1_{unity}$, and WV_{unity}) were built in such a way as to capture relevant molecular 3D-information considering molecular size, shape, and symmetry and atom distribution with respect to invariant reference frames. These descriptors are invariant to translation due to the cantering of the atomic coordinates and invariant to rotation due to the uniqueness of the principal axes, thus resulting free from the prior alignment of molecules. A fundamental role in the WHIM descriptor calculation is played by the eigenvalues and the weighted covariance matrix of the molecule atomic coordinates. Each eigenvalue represents a dispersion measure of the projected atoms along the considered principal axis, thus accounting for the molecular size along that principal direction. Relationships among the eigenvalues are used to describe the molecular shape. Within the WHIM approach, a molecule is seen as a configuration of points (the atoms) in the three-dimensional space defined by the Cartesian axes (x, y, and z). The charged partial surface area or CPSA descriptors (WNSA2 and RPCS) were originally engaged for use in structure-physical relationship studies to capture information about the features of molecules responsible for polar intermolecular interactions. The set of CPSA descriptors is based on the partial atomic charges and the partial surface area of each atom. The atom typed electrotopological state or ATES descriptor (LipoAffI) is numerical values computed for each atom in a molecule and which encode information about both the topological [38,39] environment of that atom and the electronic interactions due to all other atoms in the molecule. The electronic aspect is based on an intrinsic state plus perturbation due to intrinsic state differences between atoms in the molecule. The lipoaffinity index [41], LipoAffI is obtained by the equation of $LA=\Sigma(a_iS_i)$, where $i\neq N,O$, in addition, a_i is the coefficient and S_i is the E-state index. Thus, LA refers to the contribution to $LogP_{oct}$ from atoms other than nitrogen and oxygen. It includes contributions of carbon atoms and atoms of low hydrogen-bonding capacity, such as Cl and F as well as atoms of high polarizability such as P, S, Br, and I. The coefficient a_i can be considered as scaling factors or weights that make the contributions to LA from different atom types compatible. An extended topochemical atom or ETA descriptor (ETA_EtaP) is sufficiently rich in chemical information to encode the structural features contributing to the toxicities, and these indices may be used in combination with other topological and physicochemical descriptors for the development of predictive QSAR models. Quantum-chemically derived or QM descriptors (E_{HOMO}, E_{LUMO},

Table 3: Statistical results of training and test set performed on MLR, PCR, PLSR, and ANN regression methods (2D-QSAR) and MIFs studied on PLS method (3D-QSAR).

Method	Training set (28)					Test set (7)		3D-grid spacing	MIFs	
	r^2	q^2_{LOO}	SPRESS	SDEC	F test	r^2_{pred}	SDEP		Electro-static	Steric
2D-QSAR										
MLR	0863	0.589	0.315	0.205	21.089	0.801	0.236	-	-	-
PCR	0.848	0.538	0.352	0.216	18.619	0.694	0.293			
PLSR	0.791	0.534	0.349	0.254	12.600	0.644	0.316	-	-	-
ANN	0.954	0.815	0.215	0.119	67.168	0.862	0.196	-	-	-
3D-QSAR (atom-based alignment)										
PLS	0.940	0.590	0.313	0.136	65.344	0.871	0.190	2.0 Å	64.23%	35.77%

MLR: Multiple linear regression, PCR: Principal component regression, PLSR: Partial least square regression, ANN: Artificial neural network, LOO: Leave one out, SDEC: Standard error of calculation, SDEP: Standard error of prediction, F test: Fischer's statistics, MIFs: Molecular interaction energy fields, 3D: Three-dimensional, 2D: Two-dimensional, QSAR: Quantitative structure-activity relationship

ΔH_{FORM}, and ΔE_{GAP}) give valuable information about the reactivity, shape, and binding properties of a complete molecule as well as of molecular fragments and substituents. Because of the large, well-defined physical information content encoded in many theoretical descriptors, their uses in the design of QSAR study are the direct characterization of their molecular structure as well as their various fragments and substituents, and the proposed mechanism of action can be directly accounted for in terms of the chemical reactivity of the compounds. Consequently, the derived QSAR models will include information regarding the nature of the intermolecular forces involved in determining the biological or other activity of the compounds.

Initial analyses of the training set (28 compounds) by MLR method gave very low cross-validated coefficient, q^2 of 0.328 and squared correlation coefficient, r^2 of 0.572. A graphical inspection and the residuals of the experimental with calculated $LogIC_{50}$ values immediately indicate that the overall fit of the molecules was satisfactory except molecule 1, which behaves as an outlier. After removal of the outlier, the final model obtained by MLR [42] showed high q^2 values of 0.589 and r^2 of 0.863. This indicates that the molecule, 1 behaves differently from other molecules with respect to both molecular structures (descriptors) and antifibrosis activity. In contrast, this molecule showed the same behavior to the other regression models (PCR, PLS, and ANN) in the 2D-QSAR analysis. The statistical results obtained by MLR, PCR, and PLS methods are summarized in Table 3, and plots of the observed versus predicted $LogIC_{50}$ values are shown in Figures 3-5, respectively. The ANN calculations were carried out using IBM SPSS20 software with ANN toolbox using MLP algorithm.

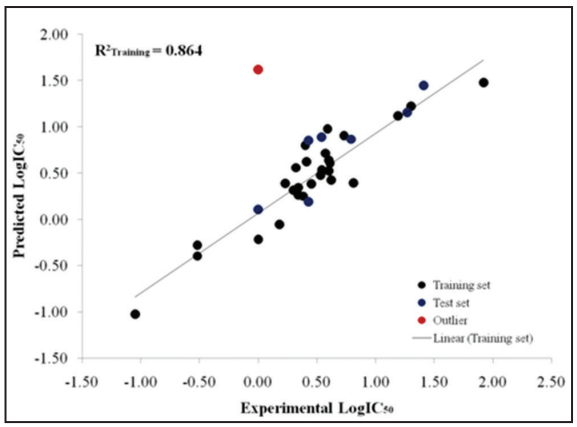


Figure 3: Graph of actual versus predicted activities for training and test set molecules by multiple linear regression method.

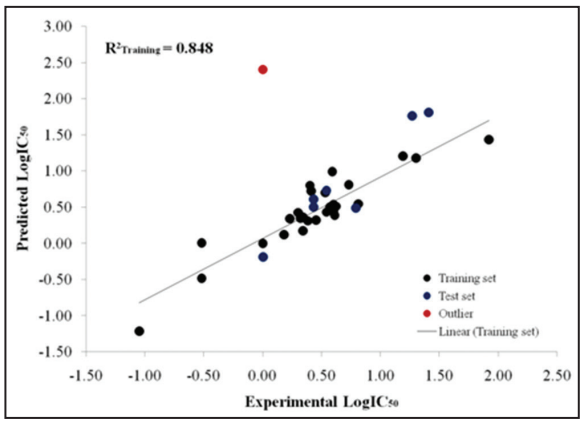


Figure 4: Graph of actual versus predicted activities for training and test set molecules by principal component regression method.

Before training process, the input (covariates) values were rescaled with standardized function. A sigmoidal

Table 4: The experimental and predicted antifibrosis activities ($LogIC_{50}$) of N_1-phenylhydroquinolinone derivatives.

ID No.	$LogIC_{50}$ (Exp)	2D-QSAR				3D-QSAR
		$LogIC_{50}$ (MLR)	$LogIC_{50}$ (PCR)	$LogIC_{50}$ (PLSR)	$LogIC_{50}$ (ANN)	$LogIC_{50}$ (MIFs)
1[x]	0.00	1.62	2.39	2.02	1.94	2.32
2	1.30	1.22	1.17	1.14	1.23	1.28
3	0.30	0.31	0.42	0.40	0.30	0.44
4*	1.27	1.15	1.76	0.96	1.29	1.34
5	0.41	0.63	0.72	0.33	0.37	0.37
6	0.62	0.42	0.51	0.49	0.73	0.51
7*	1.41	1.45	1.81	1.09	1.22	1.62
8	1.19	1.12	1.20	1.16	1.25	1.12
9	−0.52	−0.40	−0.49	−0.24	−0.77	−0.23
10*	0.43	0.85	0.50	0.57	0.49	0.60
11	0.73	0.91	0.81	0.73	0.89	0.96
12	0.54	0.53	0.43	0.64	0.54	0.62
13	0.61	0.61	0.39	0.44	0.47	0.29
14	0.45	0.38	0.32	0.31	0.40	0.37
15*	0.54	0.89	0.73	0.81	0.51	0.82
16	0.60	0.64	0.53	0.67	0.54	0.50
17	0.38	0.25	0.31	0.18	0.27	0.29
18	0.34	0.26	0.17	0.02	0.17	0.45
19	0.60	0.53	0.46	0.18	0.53	0.51
20	0.34	0.34	0.36	0.33	0.40	0.47
21	0.40	0.80	0.80	0.88	0.48	0.50
22*	0.79	0.86	0.49	0.74	0.59	0.50
23	0.57	0.72	0.50	0.36	0.45	0.58
24	0.59	0.98	0.99	0.91	0.66	0.72
25	1.92	1.48	1.43	1.59	1.84	1.74
26	0.81	0.39	0.54	0.69	0.81	0.76
27	0.18	−0.06	0.12	0.30	0.15	0.03
28	0.00	−0.22	0.00	0.17	0.16	−0.01
29*	0.43	0.19	0.61	0.29	0.70	0.49
30	−0.52	−0.28	0.00	0.21	−0.23	−0.47
31	0.53	0.48	0.70	0.88	0.46	0.54
32*	0.00	0.11	−0.19	0.62	0.34	−0.09
33	0.23	0.39	0.34	0.36	0.32	0.19
34	0.32	0.56	0.35	0.30	0.45	0.54
35	−1.05	−1.03	−1.21	−1.11	−0.89	−1.19

$LogIC_{50}$ in tenth based. [x]Outlier, *Test set. MLR: Multiple linear regression, PCR: Principal component regression, PLSR: Partial least square regression, ANN: Artificial neural network, 3D: Three-dimensional, QSAR: Quantitative structure-activity relationship

transfer function and descent gradient with momentum were designed to predict the biological antifibrosis activity of N_1-phenylhydroquinolinone derivatives. The number of neurons in hidden layer was varied, ranging from 1 to 10, and the default values were accepted for the other parameters. The inputs and outputs for the ANN simulation were the molecular descriptors and $logIC_{50}$ values, respectively. All networks were of the three-layered typed, containing a bias neuron in each layer and a single neuron in the output layer. The statistics obtained from ANN model were n=28, input columns (descriptors)=6,

net configuration=7-4-1 (6 input nodes with 1 bias, three hidden layers with 1 bias, 1 output node), with SDEC=0.119, r^2=0.954, and q^2=0.815 for training r^2_{pred} and =0.862 and SDEP = 0.196 for the test set. Although using the same descriptors for the MLR model, the ANN treatment appeared to be a significant improvement in the calculation [43]. In addition, the architecture 7-4-1 of three-layer ANN shown in Figure 6 was used to calculate the biological activities with the help of a set of molecular descriptors and experimental $LogIC_{50}$ activity.

The actual and predicted activities obtained by ANN analysis of the training and test set of compounds are shown in Table 4. Figure 7 depicts the graph plotted between actual and predicted activities of training and test set obtained by ANN. The avoidance of overfitting and overtraining is an important factor for the improvement of generalization ability in neural network studies [44]. To this point, the total data set of 35 compounds was divided into a training set (28) and test set (7).

The model generated by ANN was internally validated by a LOO method with the high value of q^2,and the model was also externally validated with r^2_{pred} calculation. Compared with the MLR, PCR, and PLSR analysis, the improved predictive performance was observed through ANN approaches. The statistical results obtained by MLR, PCR, PLS, and ANN methods are summarized in Table 3, and plots of the observed versus predicted $LogIC_{50}$ values are shown in Figures 3-5 and 7, respectively. It is obvious to see from the figures that the predicted biological activity of the PLSR model is poorer in comparison with the ANN model. The model obtained with ANN has a better correlation with experimental values than obtained with MLR, PCR, and PLSR and overall data (Table 3) reveals that the proposed models by MLR, PCR, PLSR, and ANN have high fitting and high predictive ability with lower errors of calculation.

5.3. Development of 3D-QSAR (MIFs) Models

The MIFs calculation was performed on an open3DQSAR tool [45] using the partial least square (PLS) technique [46] through the NIPALS algorithm methodology [47]. To obtain the 3D-QSAR models, PLS analysis was performed using both steric and electrostatic fields on a training set at 2.0 Å of 3D grid spacing. Energies lower than −40.0 kcal/mol and greater than 40.0 kcal/mol were cut off because a few high values in the data set may severely bias the model. Before validation, the model was selected based on the highest square correlation coefficient (r^2), and the smallest SDEC values, and the model was also validated internally and externally on the highest values of q^2 and r^2_{pred} respectively. To obtain the 3D-QSAR model, PLS analysis was performed using combined steric and electrostatic fields at 2.0Å 3D grid spacing (Table 3).

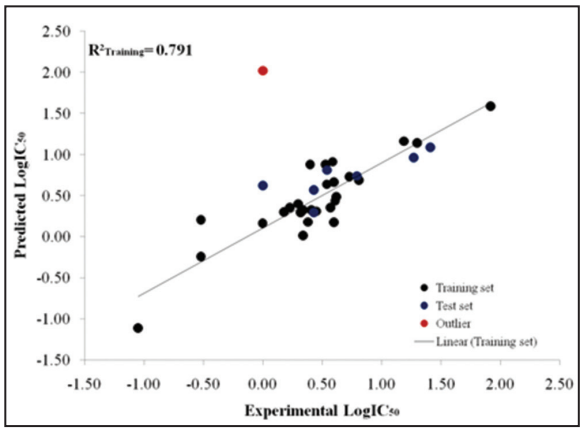


Figure 5: Graph of actual versus predicted activities for training and test set molecules by partial least square method.

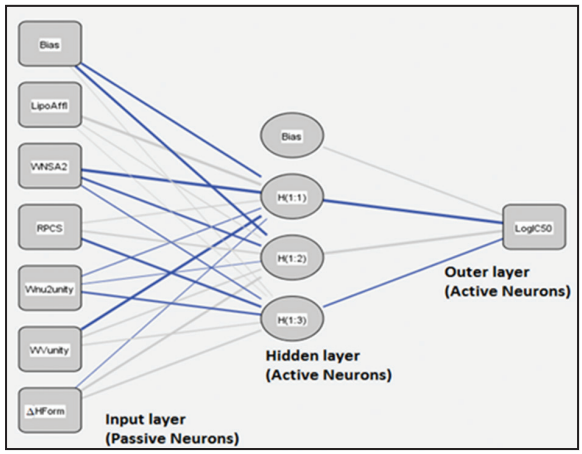


Figure 6: Architecture (7-4-1) of three layer artificial neural networks artificial neural networks.

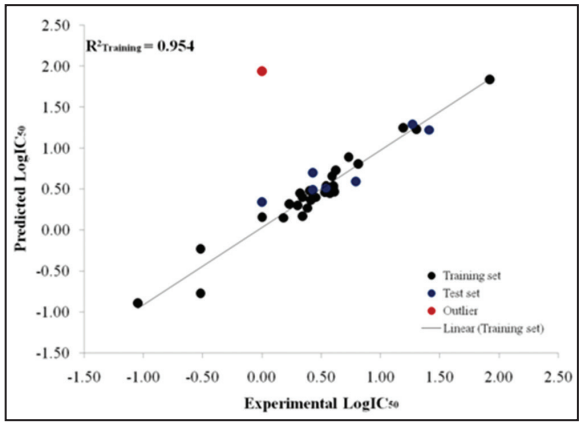


Figure 7: Graph of actual versus predicted activities for training and test set molecules by artificial neural networks method.

5.4. Evaluation of the 3D-QSAR Models

A data set of 35 N_1-phenylhydroquinolinone derivatives was divided into a training set of 28 compounds

for developing the MIFs model, and a test set of 7 compounds for evaluating the predictive ability of the models. The initial cross-validated fit on all 28 molecules was characterized by a q^2 of 0.221 (with 5 components). The residuals and graphical inspection of the experimental versus predicted ($LogIC_{50}$) values immediately indicated that the overall fit of the molecule was satisfactory except the molecule, 1 which behaves as an outlier. This particular orientation could play another specific role in the activity. Due to its different property from the other compounds, so this molecule had to be omitted from further studies. After removing the outlier, the new MIFs model was obtained with q^2 values of 0.690 (using both steric and electrostatic fields at a 2.0 Å grid spacing with 5 components). The LOO cross-validated PLS analysis results in a q^2 of 0.590, SPRESS of 0.313 using five principal components, and the noncross-validated PLS analysis yields a higher r^2 of 0.940 with a low SDEC of 0.136 and high F-value of 65.344. The model was externally validated by high r^2_{pred} of 0.871 with the low standard error of prediction (SDEP) of 0.190 for the test set. The steric and electrostatic field contributions to the model were 64.23% and 35.77%, respectively. The experimental and calculated antifibrosis activities ($LogIC_{50}$) using the best 3D-QSAR MIFs model for training and test (indicated by * marks) set are shown in Table 4. Figure 8 showed a plot of observed versus calculated antifibrosis activity of training and test set molecules through MIFs analysis [48]. Hence, the derived model was satisfactory with respect to high statistical results and predictive ability.

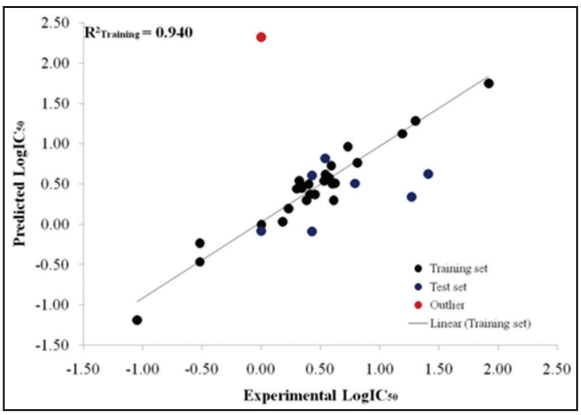


Figure 8: Graph of actual versus predicted activities by partial least square technique through molecular interaction energy fields analysis for training and test set molecules.

5.5. Assessment of the Contour Maps by MIFs

The PLS coefficient contour maps (contoured at 0.004 and energy range −40 to 40 kcal/mol) for the steric and electrostatic fields are shown in Figure 9a and b, respectively. Green regions indicate areas where a steric bulk is predicted to enhance the activity, whereas yellow contours indicate regions where a steric bulk is predicted to decrease activity (Figure 9a). Blue-colored regions indicate areas where electropositive groups are predicted to enhance the activity, whereas red regions represent areas where electronegative groups are predicted to favor the activity (Figure 9b). It is important to note that lower the IC_{50} value greater the activity and *vice versa.*

The large green contours located around R substituted benzene ring, indicate that the most bulky substituent is preferred to enhance the activity at this site. This was consistent with the experimental results, where compounds (R=Larger group than H), 12 (IC_{50}=3.5), and 27 (IC_{50}=1.5) showed comparatively higher activity than corresponding compounds (R=H), 11 (IC_{50}=5.4), and 26 (IC_{50}=6.5), respectively. The same effects were observed for the compounds 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 28, 29, 30, 31, 32, 33, 34, and 35. The activities of the compounds 2-10 (R=Larger group than H) with respect to the compound 1 (R=H) cannot be explained on green contours due to being an outlier of the compound 1. The yellow contour is located near the Y-position suggested that bulky group at the Y-position (Table 1) appeared to be unfavorable for the antifibrosis activity of N_1-phenylhydroquinolinone derivatives. For example, the activity of the compound 2 (R=o-Cl, X=O, Y=COOH, IC_{50}=20) is lower than that of the compound 12 (R=o-Cl, X=O, Y=H, IC_{50}=3.5). Similar effects were observed for compounds 4, 6, and 8 compared with 17, 20, and 23, respectively (Figure 8).

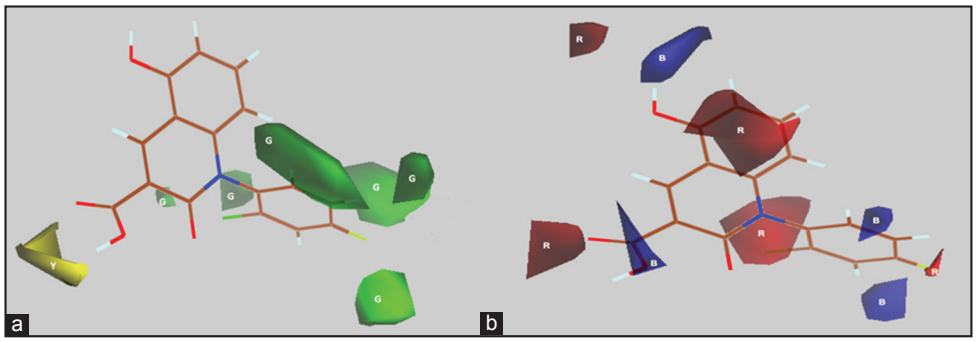


Figure 9: Partial least square coefficient maps, contoured at 0.004 and the most active compound 35 are displayed in this figure: (a) Steric contribution and (b) Electrostatic contribution (Y=Yellow, G=Green, R=Red and B=Blue).

The electrostatic contour map for MIFs model is depicted in Figure 9b. The plot shows red areas near the o- and p-positions in R-substituted benzene ring and on the X-substituted ring of the compounds. This plot accounts for the experimentally observed fact that compounds having an electron withdrawing group near this area have a higher antifibrosis activity. This is exemplified in the lower activities of compound 4 (IC_{50}=18.7, R=p-CH_3) as against a compound 3 (IC_{50}=2.0, R=p-Cl), compound 5 (IC_{50}=2.6, R=p-F), and compound 6 (IC_{50}=4.2, R=p-Br). Moreover, the better activities of compound 9 (IC_{50}=0.3) with R=3, 4-di-Cl as against compound 8 (IC_{50}=15.6) with R=2,4-di-CH_3, and also 30 (p-Br) as against 31 (p-OCH_3) appears to originate from this fact. The same reasoning seems to account for the better activity of compounds 28 and 32 as compared to 33 and also compound 24 as compared 25. The electronegative favored area (red area) of the electrostatic potential surface map (Figure 9b) over the X-substituted benzene ring is generated by the electron clouds of the benzene ring of the most active compound 35. This mapping provides the optimum electrostatic interaction of the electron cloud of the aromatic ring and additionally other red contours at o- and p-positions in R-substituted benzene ring indicates that the presence of electron withdrawing groups (2-Cl, 4-F) on these sites further increased the activity. As a result, the antifibrosis activity of the compound 35 is extremely high (IC_{50}=0.09) compared to others. It is obvious in a compound 35 (the most active compound) where the large green contours (Figure 9a) around the R-substituted benzene (2-Cl, 4-F) ring have largely improved the antifibrosis activity. On the other hand, the activity decreased by the blue area over X (-COOH) group would not be effective due to the activity increased by a red area near the same group of the compound 35. A blue contour resided in the front of a Y-group and R-substituted benzene ring, and over the X-group of N_1-phenylhydroquinilinone derivatives indicate that electropositive groups are predicted to increase activity in this area (Figure 9b). This seems to explain the better activity of compound 17 (IC_{50}=2.4) as compared to compound 19 (IC_{50}=4.0) and compound 23 (IC_{50}=3.7) as compared to compound 22 (IC_{50}=6.2). This is an indication where steric bulkiness is greatly preferred to improve the activity. The major differences of the contour maps are the appearance of green and red contours in the case of MIFs analysis. To this end, the contour similarity or dissimilarity may offer an advantage to choose the activity of the compounds.

6. CONCLUSION

We have carried out 2D- and 3D-QSAR studies of 35 N_1-substituted phenylhydroquinolinone derivatives against experimental biological activities. A set of ATES, ETA, CPSA, WHIM, and QM descriptors was used to derive 2D-QSAR models, and MIFs were used to develop the 3D-QSAR model. Our present studies have established that the model obtained from 2D-QSAR is quite reliable and pertinent to that of 3D-QSAR. We have investigated that the MIFs analyses have presented an excellent statistical result in terms of q^2 and r^2 values for N_1-substituted phenylhydroquinolinone derivatives and showed a high degree of agreement with the experimental values as well as predicted values by 2D-QSAR. Most of the compounds have potent antifibrosis activity, and the compound 35 is the most active of them, exhibited good inhibition against NIH3T3 cell proliferation. Based on the result from MIFs contour maps, it is possible to assume that X-substituted benzene ring instead of cyclohexane ring is the most important part of the activity of this series of compounds. Thus, improving the understanding of the mechanisms, cell types, and factors involved in this process is crucial to develop treatment strategies for these diseases. Assuming the QSARs analysis, we expect to find more potential compounds with the aid of the computational combinatorial chemistry method. The information provided by the QSARs analysis may give a valuable clue to design and discovery of new potential drugs.

7. ACKNOWLEDGMENT

The authors wish to thank the Authority, Department of Chemistry, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh, to use the departmental computational chemistry laboratory.

8. REFERENCES

1. P. H. McKee, J. E. Calonje, S. R. Granter, (2005) ***Pathology of the Skin,*** 3rd ed. New York: Mosby.
2. T. Maria, (2012) Mediators of fibrosis, ***Open Rheumatology Journal,*** **6:** 70-71.
3. W. A. Border, N. A. Noble, (1994) Transforming growth factor beta in tissue fibrosis, ***New England Journal of Medicine,*** **331:** 1286-1292.
4. B. M. Stramer, R. Mori, P. Martin, (2007) The inflammation-fibrosis: A jekyll and hyde role for blood cells during wound repair, ***Journal of Investigative Dermatology,*** **127:** 1009-1017.
5. S. N. Iyer, J. S. Wild, M. J. Schiedt, D. M. Hyde, S. B. Margolin, S. N. Giri, (1995) Dietary intake of pirfenidone ameliorates bleomycininduced fibrosis in hamsters, ***Journal of Laboratory and Clinical Medicine,*** **125**: 779-785.
6. J. M. Lurton, T. Trejo, A. S. Narayanan, G. Raghu, (1996) Pirfenidone inhibits the stimulatory effects of profibrotic cytokines on human lung fibroblasts ***in vitro*** (abstract), ***American Journal of Respiratory and Critical Care Medicine,*** **153:** A403.
7. K. Osanai, K. Takahashi, S. Sato, K. Iwabuchi, K. Ohtake, M. Sata, S. Yasui, (1991) Changes of lung surfactant and pressure volume curve in bleomycin-induced pulmonary fibrosis, ***Journal of Applied Physiology,*** **70:** 1300-1308.
8. S. Zia, D. M. Hyde, S. N. Giri, (1992) Development

of a bleomycin hamster model of subchronic lung fibrosis, ***Pathology,*** **24:** 155-163.
9. J. Usuki, Y. Fukuda, (1995) Evolution of three patterns of intra-alveolar fibrosis produced by bleomycin in rats, ***Pathology International,*** **45:** 552-564.
10. J. A. Jr. Cooper, D. A. White, R. A. Matthay, (1986) Drug-induced pulmonary disease. Part 1: Cytotoxic drugs, ***American Review of Respiratory Disease,*** **133:** 321-340.
11. Q. Yuan, R. Wang, Y. Peng, X. Fu, W. Wang, L. Wang, F. Zhang, Z. Peng, W. Ning, G. Hu, (2011) Fluorofenidone attenuates tubulo-interstitial fibrosis by inhibiting TGF-beta(1)-induced fibroblast activation, ***American Journal of Nephrology,*** **34:** 181-194.
12. W. Ling, L. Binu, L. Qianbin, C. Jun, T. Lijian, H. Gaoyun, (2012) Design, synthesis and anti-fibrosis activity study of n_1-substituted phenyl-hydroquinolinone derivatives, ***Molecules,*** **17:** 1373-1387.
13. A. Golbraikh, A. Tropsha, (2002) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, ***Journal of Computer-Aided Molecular Design,*** **5:** 357-369.
14. J. P. J. Stewart, (2012) ***MOPAC2012,*** Version 12.347W. Stewart Computational Chemistry.
15. J. P. J. Stewart, (2007) Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, ***Journal of Molecular Modeling,*** **13:** 1173-1213.
16. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, (2003) ***Gaussian 03W,*** Version 6.0. Pittsburgh, PA: Gaussian Inc.
17. A. D. Becke, (1993) Density-functional thermochemistry. III. The role of exact exchange, ***Journal of Chemical Physics,*** **98:** 5648-5652.
18. C. W. Yap, (2011) PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints, ***Journal of Computational Chemistry,*** **32(7):** 1466-1474.
19. N. J. Richmond, P. Willett, R. D. Clark, (2001) Alignment of three-dimensional molecules using an image recognition algorithm, ***Journal of Molecular Graphics and Modelling,*** **23:** 199-209.
20. J. Taminau, G. Thijs, H. De Winter, (2008) Pharao: Pharmacophore alignment and optimization, ***Journal of Molecular Graphics and Modelling,*** 27: 161-169.
21. M. Petitjean, (1998) Interactive maximal common 3d substructure searching with the combined SDM/RMS algorithm, ***Computers Chemistry,*** **22:** 463-465.
22. J. W. Ponder, (2004) ***TINKER: Software Tools for Molecular Design***, Version 3.0. Saint Louis, MO: Washington University of Medicine.
23. R. D. Cramer 3rd, M. Milne, (1979) ***Abstract of Papers of the American Chemical Society.*** M. 44, Computer Chemistry Section.
24. V. Centner, D. L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste, C. Sterna, (1996) Elimination of uniformatives variables for multivariate calibration, ***Analytical Chemistry,*** **68:** 3851-3858.
25. R. Grohmann, T. J. Schindler, (2008) Toward robust QSPR models: Synergistic utilization of robust regression and variable elimination, ***Journal of Computational Chemistry,*** **29:** 847-860.
26. R. D. Cramer, J. D. Bunce, D. E. Patterson, (1988) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies, ***Quantitative Structure-Activity Relationship,*** **7:** 18-25.
27. D. B. de Oliveira, A. C. Gaudio, (2001) Build QSAR: A new computer program for QSAR analysis, ***Quantitative Structure-Activity Relationship,*** **19(6):** 599-601.
28. IBM Corp. (2011) ***IBM SPSS Statistics for Windows,*** Version 20.0., NY: IBM Corp.
29. M. Smith, (1993) ***Neural Networks for Statistical Modeling,*** New York, USA: Van Nostrand Reinhold.
30. J. A. Anderson, (1995) ***An Introduction to Neural Networks,*** Cambridge, MA: MIT.
31. R. Todeschini, P. Gramatica, (1998) New 3D molecular descriptors: The WHIM theory and QSAR applications, ***Perspectives in Drug Discovery and Design,*** **9-11:** 355-380.
32. D. T. Stanton, P. C. Jurs, (1990) Development and use of charged partial surface area structural descriptors in computer assissted quantitative structure property relationship studies, ***Analytical Chemistry,*** **62:** 2323-2329.
33. H. Sklenar, J. Jager, (1979) Molecular structure-biological activity relationships on the basis of quantum-chemical calculations, ***International Journal of Quantum Chemistry,*** **16:** 467-484.
34. L. Buydens, D. Massart, P. Geerlings, (1983) Prediction of gas chromatographic retention indices with topological, physicochemical and quantum chemical parameters, ***Analytical Chemistry,*** **55:** 738-744.
35. M. Cocchi, M. C. Menziani, P. G. De Benedetti, (1992) Theoretical versus empirical molecular descriptors in monosubstituted benzenes: A chemometric study, ***Chemometrics and Intelligent Laboratory Systems,*** **14:** 209-224.
36. K. Roy, G. Ghosh, (2004) QSTR with extended topochemical atom indices. 2. Fish toxicity of substituted benzenes, ***Journal of Chemical Information and Computer Sciences,*** **44:** 559-567.
37. K. Roy, R. N. Das, (2011) On some novel extended topochemical atom (ETA) parameters for effective

encoding of chemical information and modelling of fundamental physicochemical properties, ***SAR QSAR Environmental Research,* 22:** 451-472.
38. L. H. Hall, L. B. Kier, (1995) Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information, ***Journal of Chemical Information and Computer Sciences,* 35:** 1039-1045.
39. R. Liu, H. Sun, S. S. So, (2001) Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. Blood-brain barrier penetration, ***Journal of Chemical Information and Computer Sciences,* 41:** 1623-1632.
40. P. Gramatica, M. Corradi, V. Consonni, (2000) Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by different sets of molecular descriptors, ***Chemosphere,* 41:** 763-777.
41. C. Ehrhadt, K. J. Kim, (2007) ***Drug Absorption Studies: In Situ, In Vitro and In Silico Models***, New York: Springer.
42. V. K. Vyas, M. Ghate, H. Katariya, (2011) 2D and 3D-QSAR study on 4-anilinoquinozaline derivatives as potent apoptosis inducer and efficacious anticancer agent, ***Organic and Medicinal Chemistry Letters,* 1(1):** 1-11.
43. M. A. K. Liton, U. Salma, A. C. Bhowmick, (2014) Cytotoxicity and 2D-QSAR study of some heterocyclic compounds, ***Arabian Journal of Chemistry,* 7:** 639-646.
44. I. V. Tetko, D. J. Livingstone, A. I. Luik, (1995) Neural network studies. Comparison of overfitting and overtraining, ***Journal of Chemical Information and Computer Sciences,* 35:** 826-833.
45. P. Tosco, T. Balle, (2011) An open-source software aimed at highthroughput chemometric analysis of molecular interaction fields, ***Journal of Molecular Modeling,* 17:** 201-208.
46. L. Stahle, S. Wold, (1987) Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study, ***Journal of Chemometrics,* 1:** 185-196.
47. S. Wold, M. Sjöström, L. Eriksson, (2001) PLS-regression: A basic tool of chemometrics, ***Chemometrics and Intelligent Laboratory Systems,* 58:** 109-130.
48. M. A. K. Liton, A. C. Bhowmick, M. A. Ali, (2013) 3D-QSAR MIFs studies on 3,5-substituted-1,4,2-dioxazoles derivatives using Open3DQSAR tools, ***Universal Journal of Chemistry,* 1(2):** 71-76.

***Bibliographical Sketch**

M. Abul Kashem Liton is working as Assistant Professor at the Department of Chemistry, Mawlana Bhashani Science and Technology University (MBSTU), Santosh, Tangail-1902, Bangladesh. At present, he teaches Organic Chemistry to the graduate and under graduate students in MBSTU. He is also supervising two students in M.Sc. thesis. He had completed M.Sc. degree in Chemistry (Organic Chemistry with thesis) at Jahangirnagar University, Savar, Dhaka, Bangladesh, and now, he is studying on the synthesis of heterocyclic compounds, biological activity, quantitative structure-activity relationship (QSAR) and theoretical investigation in his Ph.D. program in the Department of Chemistry at Jahangirnagar University. His research interests are Organic and Medicinal Chemistry; Synthesis and Biological studies of bioactive heterocycles targeting Anticancer, Antiviral, and Antimicrobial Studies; Computer aided molecular modeling of bioactive heterocycles; Drug design and Chemical Biology-Drug discovery. He had published more than 11 research papers in reputable national (3)and international(8) journals.